Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 15 (2011) 3227 – 3233

Procedia Engineering

www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

A Minimum Cost of Network Hardening Model Based on Attack Graphs

MA Jun-chun[a*,b], WANG Yong-jun[a], SUN Ji-yin[b], CHEN Shan[b]

[a]*School of Computer Science, NUDT, Changsha 410073, China*
[b]*Research Inst. of High-tech Hongqing Town, Xi'an 710025, China*

Abstract

In order to improve network's security, a minimum cost of network hardening model (MCNHM) based on attack graphs is presented. Firstly, the bidirectional-based search strategy is used to search the network vulnerabilities' relationship, which improves the generation efficiency of attack graphs, and reduces the system resource consumption; Secondly, this model gives the formal definition of minimum-cost of network hardening; Finally, it combines attack graphs and genetic algorithm, and transforms the problem of minimum cost of network hardening to a non-restraint optimization problem with penalty by establishing the corresponding mathematical model, which guarantees the network security with the least cost. This model is an important component of the National High-Tech Research and Development Plan of China, under Grant No.2009AA01Z432, a great of experimental results show that this model can find the minimum cost of target network, so it can help network security managers carry on safety protection in pertinence, and has important practical significance.

Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: network security; attack graphs; bidirectional-based search; genetic algorithms; minimum-cost

1. Introduction

With the rapid development of Internet and the wide application of network technology, network grows rapidly around the world. At the same time, network attack shows complex and diverse trend, so network

* Corresponding author. Tel.: +86-13484540452.
E-mail address: chenshan1223@126.com.

doi:10.1016/j.proeng.2011.08.606

security becomes a hot topic. Network vulnerability is the leading reason of network attack. Attack graphs based on network status and vulnerability information show the relationship between vulnerabilities, through which, network security managers can directly observe the attacker using a number of vulnerabilities how to attack and all the attack paths, which is easy for network security managers to guard the network. But at present, when analyzing attack graphs, the biggest challenges is how to quickly understand a large and cumbersome attack graph, and finally propose a minimum cost of network hardening. Therefore, the study on minimum cost of network hardening has an important and practical significance.

Now there is a variety of methods of constructing attack graphs[1-5] and seeking the minimum cost of network hardening. By analyzing these existed methods, there have two deficiencies as follows: firstly, attack graphs construction's problems are high complexity, low scalability and state explosion, which lead large consumption of resources when constructing attack graphs; secondly, the study on seeking the minimum cost of network hardening is only at the level of analysis, which doesn't consider implementation issues and how to apply to the real network environment.

In order to solve above problems, first of all, using the method proposed in literature [6] to quickly construct attack graphs, it combines the forward attack graphs construction algorithm and the reverse attack graphs construction algorithm, using multi-thread approach at the same time. Secondly, based on the constructed attack graphs, using genetic algorithms, this paper transforms the problem of the minimum cost of network hardening into a problem of non- constrained optimization with penalty, which ensures the security of target network with minimum cost.

2. Formal definition of minimum cost of network hardening

On the one hand, the minimum cost of network hardening is a NP complete problem[7], for which it is impossible when using exact search method, but to use heuristic search method; On the other hand, compared with other optimization algorithms, the genetic algorithm has the following advantages: (1)large coverage benefiting the global merit; (2)less dependence on the problem and stronger robustness in solution; (3)characteristics of parallel computing to improve the speed of calculation; (4)especially for solving the large complex system optimization problem, and so on. Therefore, this paper presents a MCNHM which combines the genetic algorithm and attack graphs.

Definition 1: Using attack graphs, network security administrators can identify the relationship between vulnerabilities and the resulting potential threats in target network, therefore they can take safety measures to guard the network, which controls the residual risk in tolerance range. Because different security measures need different costs, in a premise of limited resources, it is advisable to use the minimum cost, which is the minimum cost of network hardening.

Definition 2: Based on attack graphs, the MCNHM is defined as the following 8 per group: AGGA = (A,B,N,F,S,C,M,E).

(1) A shows attack graphs constructed by algorithms proposed in literature[6];

(2) B is the process of encoding chromosome, which decides the initial population. The encoding basic requirement is to establish correspondence between chromosome's genotype (code string) and phenotype (parameters). This paper takes the binary encode, and the corresponding relationship is the conversion between binary and decimal;

(3) N is the initial population. Population size N should be made affordable. If N is too small, it is difficult to obtain the global minimum cost. If N is too large, then the speed of converge is slower. It is advisable when N is two times than code length.

(4) F is the fitness function of chromosomes. Fitness function is the basis of genetic algorithm selection, crossover and mutation operations, the selection of which is a complex problem with high skill and less sufficiency in theoretical research.

(5) S is one of the basic genetic operators: selection. From the initial population, according to certain standards, we select parental chromosome. There is a variety of selection methods, and the usual methods are expected value method, fitness probability method (roulette wheel method), and the best individual preservation. According to the application, we use expected value method.

(6) C is one of the basic genetic operators: crossover. It is an exchange of the genes of two chromosomes to generate new offspring. Crossover operation can make changes in gene sequencing more flexible and achieve a real sense of human recombinant, which is suitable for large-scale operation. The usual crossover methods are single point crossover, double point crossover, sequential crossover and position crossover, etc.. Depending on the application, we use single point crossover.

(7) M is one of the basic genetic operators: mutation. It is a change of a gene or several genes in the organisms, but such probability is very small. In natural selection, it is usually a few thousandths or a few percent. As this paper takes binary encoding, the so-called mutation is to make the value of gene contrary, for changing 0 to 1 and changing 1 to 0.

(8) E is the terminal conditions of genetic algorithm. After selection, crossover and mutation operator, we obtain new populations' genotypes, and take fitness calculation to them, this process is called iteration. According to a large number of repeated experiment results, we summed up the terminal conditions of genetic algorithm:

Firstly, according to the number of initial attributes, namely, the size of network, the user can set the minimum number of iterations. For small scale network, the number of iterations is smaller, such as 10, but for large-scale network, it should be bigger, such as 100.

Secondly, when the number of iterations exceeds the minimum number of iterations, it is advisable to determine whether there is a selection probability of chromosome is more than 80% (user can set by himself). If existed, then the program ends; if all the chromosomes' probabilities in the new generation are same, then the program also ends.

3. MCNHM based on attack graphs

3.1. Method of constructing attack graphs

Literature [19] presented a scalable, bidirectional-based search strategy to construct attack graphs. On the one hand, it combined the forward attack graphs construction algorithm and the reverse attack graphs construction algorithm. The search speed was increased greatly, and the algorithm's time complexity was reduced. On the other hand, it modeled the target network in four levels: hosts' accessibility, security systems, host systems and network services, suggested a technology that can automatically acquire the parameters of hosts' accessibility, so it effectively supported us to model a large-scale target network automatically and reduced the algorithm's space complexity at the same time.

3.2. Model implementation

After attack graphs are generated in the previous section, based on logical reasoning, according to attack graphs, this paper firstly translated attack goal into a Boolean expression g(x), secondly, found the minimum cost of network hardening. In practical applications, when finding minimum cost of network hardening, it must meet g(x)=0, then from the set {x1, x2, x3 ... xn} which meets g(x)=0 to select the minimum cost of x, so the problem of finding the minimum cost of network hardening is a constrained

optimization problem. To solve this problem, we use a penalty method, whose basic idea is to give corresponding punishment to individual who violates constraints, and in fitness function penalty function shows this penalty. So the constrained optimization problem is translated into a non-constrained optimization problem with penalty. The flow of MCNHM's algorithm of this model is shown in Fig.1.

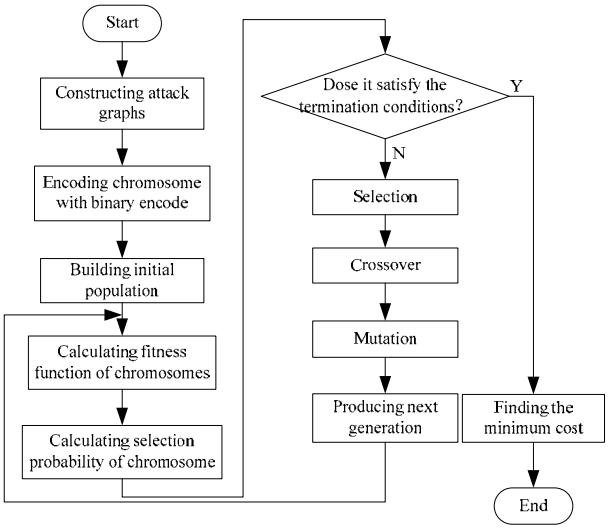

Fig. 1. The flow of MCNHM's algorithm

4. Analysis of a network instance

4.1. Experimental conditions

Fig.2 shows an enterprise local area network instance. It has two servers, respectively labeled as 1 and 2. The server 1's operating system is Linux Kernel 2.6.21, and the server 2's is Linux Kernel 2.6.13. server 2 has a weak point that can be attacked by local memory overflow, so the attacker can obtain the server's root access. The host service runs Apache 2.0.46 of which there is weak point that can be attacked by remote memory overflow to obtain user access. The firewall divides entire network into the internal network and the external network, it allows external hosts to access internal server 1 and 2. It is assumed that the attacker initially can access to user 0 and server 1, and he wants to obtain root access to server 2. Table 1 shows the network's vulnerability.

The attacker's goal is to obtain root access to server 2, while it has vulnerability CVE-2005-2558 and CVE-2006-2451, so he has two ways to obtain user access: one is from host 0, it can be attacked by remote memory overflow, the other is from host 0, it can be attacked by local memory overflow. When the attacker gains user access to server 2, he can use the local memory overflow attack to achieve his target. Fig.3 shows the corresponding attack graph, in which user(0) shows that the attacker has user access to host 0; root (2) shows that the attacker has root access to server 2, runProcess (2,Apache) shows that server 2 runs the Apache service, Rbof (1,2,Apache) shows that from server 1, server 2 can be attacked by local memory overflow, Rbof (0,2,Apache) shows that from host 0, server 2 can be attacked by remote memory overflow, Lbof (2,Linux) shows that server 2 can be attacked by local memory overflow.

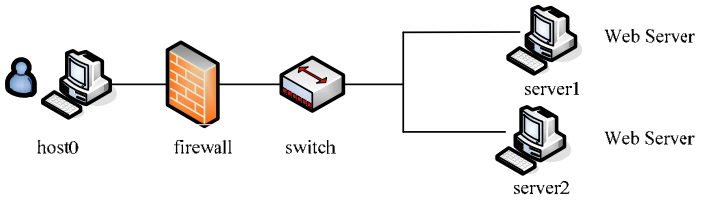

Fig. 2. Experiment network's topology

In order to better describe MCNHM, without loss of generality, we abstract Fig.4 which generated by literature [19] to Fig.3, and assume Cost(c_1)=10, Cost(c_2)=1, Cost(c_3)=15, Cost(c_4)=1, the attacker's target is {c_7}, then seek the network instance's minimum cost of network hardening.

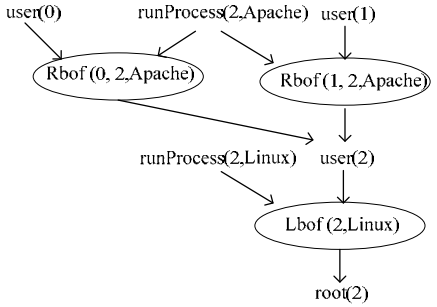

Fig. 3. Experiment network's attack graph

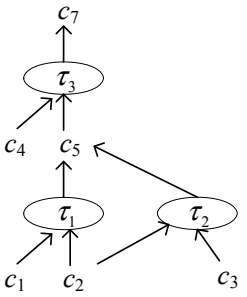

Fig. 4. Attack graph's abstract graph

4.2. Experimental results and analysis

4.2.1. Manual analysis of the minimum cost of network hardening

In order to validate the proposed model, we need to analyze minimum cost of network instance hardening by manual. From Fig.5 we can obtain all attack paths to reach target c_7: $\tau_1 \to \tau_3$ and $\tau_2 \to \tau_3$. In order to ensure the security of target c_7, because of attribute c_5 can be reached from τ_1 and τ_2, therefore, the relationship of atomic attack τ_1、τ_2 is "or", that is, as long as one of them is used, then the attacker can obtain attribute c_5 and so on. The best way to prevent atomic attack is to eliminate the premise attributes $c_1c_2c_3c_4$ which cause the atomic attack. Through analysis, we can obtain the network instance hardening is c_1,c_2,c_3,c_4; c_1,c_2,c_4; c_2; c_2,c_4 and so on, based on the assumption Cost(c_1)=10, Cost(c_2)=1, Cost(c_3)=15, Cost(c_4)=1, we can obtain the minimum cost of network instance hardening is c_2.

4.2.2. Result of MCNHM

According to the proposed model and its flow chart, we need keep to the following steps:

Step1: The process of encoding chromosomes. In Fig.5, there are four initial attributes $c_1c_2c_3c_4$. we take binary encoding, in which 0 shows that we don't need to operate the attribute and 1 shows that we need to remedy it. For example, 0001 shows that we only need to remedy c4. Since the chromosome genotypes are between 0001 and 1111, the corresponding decimal values are between 1 and 15.

Step2: Determine the initial population. Because the chromosome consists of four genes and the length of encoding is 4, we take N=8.

Step3: Determine the fitness function of t_i. According to Fig.5, we need translate the target c_7 into Boolean expression g(x); then transform the problem of minimum cost of network hardening into a problem of non- constrained optimization with penalty. Through a large number of experimental data, the fitness function for c_7 when r=30, the punishment is appropriate.

$$f_i = -[\cos t(c_1)\cdot c_1 + \cos t(c_2)\cdot c_2 + \cos t(c_3)\cdot c_3 + \cos t(c_4)\cdot c_4] + r[1 - g(x)]$$

$$t_i = (f_i - f_{\min} + 1)\cdot 100$$

$$g(x) = c_4 \wedge ((c_1 \wedge c_2) \vee (c_2 \wedge c_3))$$

Step4: Calculate the probability of selection $P_i = t_i / \Sigma t_i$.

Step5: Determine whether meeting the terminal conditions. If an individual selection probability P_i>80%, or in populations the same individuals' proportion>80%, or the number of iterations>300, then the iteration terminates, and the step9 will be carried to find the optimal solution. Otherwise, the iteration will be continued.

Step6: Selection operation. According to expected value method, it needs to select an individual as a male parent each time. The individual selection probability P_i is greater, and it becomes male parent's probability is greater.

Step7: Crossover operation. Firstly, match the male parent randomly, secondly, for each pair of male parent, identify the crossover starting position a randomly, then exchange chromosome after a.

Step8: Mutation operation. After crossover operation, we obtain new individuals, in probability of 1% we carry mutation operation on them, that is, when an individual chromosome varies, we should take opposite of it.

Step9: Iteration of step4.

Step10: The output of MCNHM is the optimal solution c_2, that is, c_2 is the minimum cost of network instance hardening.

4.3. Analysis of algorithm complexity

In order to verify the feasibility and effectiveness of MCNHM, and whether it can be used in large-scale complex networks, this paper respectively carries experiments when the number of initial attributes is 50, 100, 200, 500, 1000 and 2000. at the same time the average time of the algorithm is recorded. The experimental results are shown in Table 1, the hardware platform is P3 2.4GHz processor, 2G memory. Fig.5 shows the experimental curve.

Table 1. Calculation experiment of minimum cost

initial number of attributes	50	100	200	500	1000	2000
average number of iterations	554	926	6543	7000	7000	7000
average time /ms	98	317.50	1236.92	10500.00	56407.00	287062.00
average time of a single iteration /ms	0.1769	0.3429	0.1890	1.5000	8.0581	41.0089

It can be seen from the experimental results, with the number of hosts increases in the network, the initial attributes also increase, so when using MCNHM to seek minimum cost of network hardening, the average number of iterations and the average calculation time also increase, but the rate of increase is linearly, and when the initial number of attributes is 500, the average calculation time is merely18s, so the model can be applied to large and complex networks.

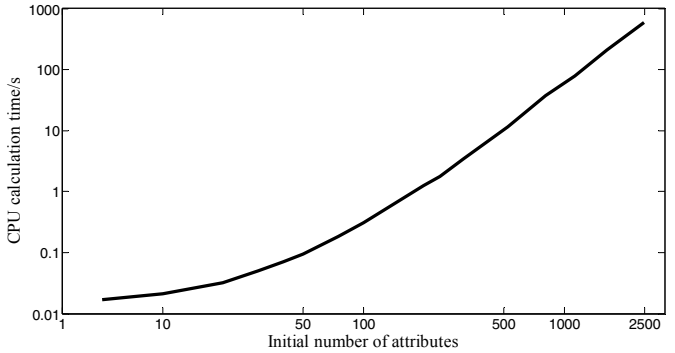

Fig. 5. Experimental curve of MCNHM

5. Conclusion

Because the minimum cost of network hardening is a NP-complete problem, based on attack graphs, a MCNHM is proposed. This paper has the following improvements: a) use Mysql database, Visual C + +, Graphviz, Visio 2007 and other tools to construct attack graphs, which effectively reduces the algorithm's complexity. b) The genetic algorithm is applied to MCNHM, because it has the characteristics of parallel computing, especially for complex system optimization problems, a great deal of experimental results show that it can be applied to large and complex network systems, and has good scalability.

Acknowledgements

This research is supported by the National High-Tech Research and Development Plan of China, under Grant No.2009AA01Z432; the National Natural Science Foundation of China, under Grant No. 60873215.

References

[1] S. Jajodia, S. Noel, and B. O'Berry. Topological Analysis of Network Attack Vulnerability[M]. Kluwer Academic Publisher, 2003.

[2] X. Ou, S. Govindavajhala, and A. Appel. MulVAL: A logicbased network security analyzer[J]. In Proceedings of the 14th USENIX Security Symposium, pages 113–128, 2005.

[3] R.P. Lippmann K.W. Ingols, et al. Evaluating and Strengthening Enterprise Network Security Using Attack Graphs. Project Report, ESC-TR-2005-064

[4] Kyle Ingols, Richard Lippmann, Keith Piwowarski. Practical Attack Graph Generation for Network Defense[J], MIT Lincoln Laboratory 2006

[5] Leevar Williams, Richard Lippmann, and Kyle Ingols. GARNET: A Graphical Attack Graph and Reachability Network Evaluation Tool[J]. Springer-Verlag Berlin Heidelberg pp. 44–59, 2008.

[6] Ma Jun-chun, Wang Yong-jun, Sun Ji-yin. A Scalable, Bidirectional-Based Search Strategy to Generating Attack Graphs. In: Proc. of 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010), 2010.2976-2981.

[7] CHEN Feng, ZHANG Yi, SU Jin-shu, HAN Wen-bao. Two formal analyses of attack graphs. Journal of Software[J], 2010,21(4):838-848.